Autosomal dominant polycystic kidney disease

Spithoven, Edwin

6

Estimation of total kidney volume in Autosomal Dominant Polycystic Kidney Disease

Edwin M. Spithoven
Maatje D.A. van Gastel
A. Lianne Messchendorp
Niek F. Casteleijn
Joost P.H. Drenth
Carlo A.J.M. Gaillard
Johan W. de Fijter
Esther Meijer
Dorien J.M. Peters
Peter Kappert
Remco J. Renken
Folkert W. Visser
Jack Wetzels
Bob Zietse
Ron T. Gansevoort

On behalf of the DIPAK Consortium

Submitted

ABSTRACT

Introduction

In Autosomal Dominant Polycystic Kidney Disease (ADPKD), total kidney volume (TKV) predicts kidney function decline. To alleviate the time-consuming measurement of TKV (mTKV), two methods have been proposed to estimate TKV ($eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$). The aim of this study was to validate these estimation methods in ADPKD patients with a wide range of kidney function.

Methods

First, we investigated the reproducibility of mTKV and eTKV in a test-set of ADPKD patients. Second, we assessed bias, precision and accuracy of eTKV in a cohort of ADPKD patients. Third, in a subgroup we determined the association between change in mTKV and change in eTKV over time.

Results

In the test-set of 10 ADPKD patients, intra- and inter-coefficients of variation were low for mTKV (1.8 and 2.3%), $eTKV_{ELLIPSOID}$ (3.9 and 6.3%) and $eTKV_{PANK}$ (3.0 and 3.4%). In the cohort of 264 ADPKD patients (50.0% men; mean age 46.7 ± 8.8 yr, eGFR 57.0, range 15.2 to 129.2 ml/min*1.73m^2) mTKV was 1.87 (1.24 - 2.67) L and $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$ 1.93 (1.25 - 2.80) L and 1.80 (1.17 - 2.52) L, resp. Bias was -0.03 ± 3.4% for repeat mTKV and 1.2 ± 9.3% and 4.8 ± 8.3% for $eTKV_{ELLIPSOID}$ (p=0.7) and $eTKV_{PANK}$ (p=0.002), resp. MRI T1-weighted compared to T2-weighted images showed significant higher bias for $eTKV_{ELLIPSOID}$ (p=0.036). When stratified for kidney function, $eTKV_{PANK}$ was less accurate than $eTKV_{ELLIPSOID}$ at eGFR <60 ml/min/1.73m^2. When using eTKV instead of mTKV, 92% of the patients were classified in their corresponding risk class. In addition, change in mTKV during three years of follow-up in 48 patients correlated well with change in eTKV. No significant differences were observed between % change in mTKV (16.7 ± 17.1%) and % change in $eTKV_{ELLIPSOID}$ (19.3 ± 16.1%) and $eTKV_{PANK}$ (17.8 ± 16.1%). Analysis time per MRI was 55, 15 and 5 minutes for mTKV, $eTKV_{PANK}$ and $eTKV_{ELLIPSOID}$, resp.

Conclusion

We showed that both TKV estimation methods perform relatively well, and can accurately detect change in TKV over time. Since measurement of $eTKV_{ELLIPSOID}$ is more feasible to estimate TKV, we advise to use this method to estimate TKV in clinical care.

INTRODUCTION

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is characterized by the formation and growth of numerous cysts in both kidneys, leading to an increase in kidney volume. These cysts compress healthy kidney tissue, causing progressive kidney function decline, and in most patients ultimately a need for renal replacement therapy. In ADPKD patients, total kidney volume (TKV) has been shown to be an early marker of disease severity and to predict kidney function decline.[1] Measurement of TKV is therefore used to assess prognosis in clinical care and for selection of patients for randomized controlled trials.[2] In these trials that investigate potential treatments for ADPKD patients assessment of TKV is often used as primary or secondary study endpoint.[3-5]

The gold standard method to measure TKV is the stereologic method. Computer Tomogram or Magnetic Resonance Images (MRI) is used, in which per slice the kidney boundaries are traced manually using dedicated software. TKV is calculated from a set of contiguous images by summing the products of the area measurements within the kidney boundaries and slice thickness.[6] This method is laborious, which limits its use in trial settings, but especially in clinical care.

In case kidney volume could be estimated with sufficient accuracy and reliability, it would alleviate the time consuming process of kidney volume measurement. Recently, two kidney volume estimation methods have been developed: the mid-slice method by the CRISP consortium[7] and the ellipsoid method by the Mayo Clinic.[2] For both methods, measured and estimated kidney volumes appeared to be well correlated, but other groups have yet not validated these methods. In addition, the mid-slice method was developed in a cohort that included only patients with a creatinine clearance >70 ml/min. Such patients have in general relatively small kidneys making stereologic measurement of TKV relatively easy, which may have influenced the results that were obtained. This method should therefore also be validated in patients with lower kidney function. Estimation methods to assess TKV may also be used in clinical trials, but only when they can accurately and reliably detect changes in TKV over time. To our knowledge these issues have not been investigated yet.

Given these considerations, the objective of the present study was to validate cross-sectionally the aforementioned methods to estimate TKV in a patient group with a wide range of kidney function. Furthermore, we investigated in a longitudinal study whether these estimation methods can accurately detect changes in TKV.

METHODS AND MATERIALS

Patients and study design

For this study, all MRI images of ADPKD patients that were available per Augustus 1, 2014, were used. These patients participated in one of three studies that were performed by the departments of Nephrology at the University Medical Centers (UMCs) of Groningen, Leiden, Nijmegen and Rotterdam (all in The Netherlands). Details of the study protocols have been published elsewhere.[4,8,9] Subjects were diagnosed with ADPKD based on the modified Ravine criteria.[10] The Medical Ethical Committee of the UMC Groningen approved the protocols of the three studies that were conducted in accordance with the International Conference of Harmonization Good Clinical Practice Guidelines and in adherence to the ethical principles that have their origin in the Declaration of Helsinki. All patients gave written informed consent.

Measurement and collections

All participants collected a 24h urine sample the day preceding the MRI-scan, in which urinary albumin concentration was measured. At the outpatient clinic on the day of the MRI, blood pressure was assessed at rest in a supine position with an automatic device (Dinamap; GE Medical Systems) for 15 minutes and weight and height were determined. Blood samples were drawn for determination of creatinine with an enzymatic assay (isotope-dilution mass spectrometry traceable; Modular, Roche Diagnostics), which was used to estimate glomerular filtration rate was using the CKD-EPI equation ($eGFR_{CKD\text{-}EPI}$).[11,12]

Magnetic resonance imaging

All participants underwent a standardised abdominal MRI protocol without the use of intravenous contrast to measure TKV. For the specific MRI-protocol, see the Supplementary Methods.

Gold standard method; measured total kidney volume (mTKV)

Kidney and liver volumes were measured preferably on the coronal T2-single shot fast spin echo sequence. The kidney and liver volumes were measured using the stereologic method. The kidney and liver boundaries were manually traced using the commercially available software Analyze Direct 11.0 (Analyze Direct, Inc., Overland Park, KS, USA). The kidney and liver volumes were calculated from the set of contiguous images by summing the products of the area measurements within the kidney or liver boundaries and slice thickness.[6] Non-renal parenchyma e.g. the renal hilus, was excluded from measurement.

Estimation methods; estimated TKV (eTKV)

The two formulas used to estimate kidney volume were derived from literature.[2,7] We first used the mid-slice method to estimate TKV ($eTKV_{PANK}$).[7] The mid-slices of the coronal MR images were selected for each kidney separately. The mid-slice was defined as the slice of which the slice number corresponds to the half of the sum of the numbers of the first and the last slice that contained the kidney. If the sum was odd, the mid-slice number was rounded up. $eTKV_{PANK}$ was calculated in mL, with mid-slice area and slice thickness in mm^2 and mm, respectively. $eTKV_{PANK}$ was calculated as the sum of the left eKV_{PANK} and right eKV_{PANK}, with left eKV_{PANK} = 0.624* mid-slice area* number of slices covering the left kidney * slice thickness/1000, and right eKV_{PANK}) = 0.637 * mid-slice area * number of slices covering the right kidney* slice thickness/1000.

Second, we used the ellipsoid method to estimate TKV ($eTKV_{ELLIPSOID}$).[2] Per kidney, length was measured as the average maximal longitudinal diameter measured in the coronal and sagittal plane. Width was obtained from the transversal image at maximum transversal diameter, and depth was measured from the same image perpendicular to the width measurement. $eTKV_{ELLIPSOID}$ was calculated in mL, with length, width and depth in mm, respectively. $eTKV_{ELLIPSOID}$ was calculated as the sum of the left $KV_{ELLIPSOID}$ and right $KV_{ELLIPSOID}$, both derived by the equation $\pi/6$ * ($length_{coronal}$ + $length_{sagittal}$)/2 * width * depth/1000. Of note, to assess $eTKV_{ELLIPSOID}$ no specific software is necessary, in contrast to measurement of mTKV and $eTKV_{PANK}$.

Statistical analyses

All analyses were performed with SPSS, version 22.0 (SPSS Inc). Normality of data was assessed by drawing Q-Q plots. Normal distributed variables are expressed as mean ± standard deviation (SD), whereas non-normal distributed variables are given as median with interquartile range (IQR). Baseline characteristics of the study population are given overall (Table 1) and stratified for eGFR < and ≥ 60 mL/min/1.73m^2 (Supplementary Table 1).

Table 1. Participants' characteristics

	Whole study group	Patients with follow-up
N	264	48
Age (y)	46.7 ± 8.8	39.2 ± 7.4
Male (%)	50.0	70.8
Body mass index (kg/m^2)	26.7 ± 4.3	26.3 ± 3.4
Body surface area (m^2)	2.0 ± 0.2	2.1 ± 0.2
Diastolic blood pressure (mm Hg)	81.5 ± 9.5	82.6 ± 8.8
Systolic blood pressure (mm Hg)	131.8 ± 13.0	132.9 ± 11.6
Antihypertensive medication (%)	86.5	81.3
Plasma creatinine (mmol/l)	126.0 ± 42.2	102.1 ± 31.7
eGFR (ml/min/1.73m^2)	57.0 ± 21.1	79.7 ± 22.6
24h Urine volume (L)	2.38 ± 0.77	2.48 ± 0.87
Albuminuria (mg/24h)	45.8 (21.5 - 88.7)	46.2 (19.0 - 181.0)
Total kidney volume (L)	1.87 (1.24 - 2.67)	1.79 (1.36 - 2.56)
Left kidney volume (L)	0.98 (0.63 - 1.44)	0.99 (0.73 - 1.39)
Right kidney volume (L)	0.87 (0.58 - 1.34)	0.80 (0.57 - 1.17)

Abbreviations: eGFR, estimated glomerular filtration rate (CKD-EPI equation).

Differences between groups were tested using a two-sample *t*-test for normal distributed and a Mann-Whitney U test for non-normal distributed data. For paired analyses, a paired *t*-test was used for normal distributed and a Wilcoxon-signed rank test for non-normal distributed data. McNemar test was used for paired nominal data. A two-sided $p<0.05$ was considered to indicate statistical significance.

In a test-set of ten patients, stratified for kidney volume and MRI-scanner, kidney volumes were measured and estimated twice by four reviewers. All reviewers were blinded for their previous measurement results. These results were analysed to calculate intra- and inter-observer coefficients of variation (CV).

To investigate whether eTKV correlated with mTKV, orthogonal regression analysis was performed, and Lins' concordance correlation coefficient (CCC) was calculated using all MRI-scans of our cohort. Agreement between eTKV and mTKV was evaluated by Bland-Altman analyses, with calculation of agreement limits (95% Confidence Interval [CI]). Performance of the estimation methods compared with mTKV was assessed using bias, precision and accuracy. For cross-sectional analyses, bias is expressed as mean percentage difference (mTKV - eTKV)/ mTKV*100%, with positive values indicating an underestimation of mTKV. Precision was defined as 1 standard deviation of bias. Accuracy was calculated as the percentage of

eTKV values within 10%, 15% and 20% of mTKV [P_{10}, P_{15}, and P_{20} respectively]). To investigate whether bias is dependent on patient or MRI characteristics, we performed regression analyses between bias and various variables i.e. age, length, BMI, liver volume and T1/T2 weighted images in univariate analyses. Differences in bias among the various scanners that were used, were tested with analysis of variance (ANOVA). As standard quality control 10% of all MRI-scans were measured twice for mTKV. This was done to ensure that the observers maintained low inter-observer variability. These scans were used to assess precision and bias of mTKV ($mTKV_{REPEAT}$).

To investigate whether the estimation methods can accurately detect changes in TKV, data of patients who had follow-up MRIs available were used. For these longitudinal analyses bias is expressed as (%change in mTKV - %change in eTKV). Importantly, all follow-up scans were performed at the same MRI-scanner as at baseline, and TKV was measured and estimated using the same images series as at baseline, by reviewers blinded for the baseline measurement results.

To assess the consequences of using eTKV instead of mTKV, two analyses were performed. First, the effect on classification based on disease prognosis was assessed. To assess prognosis for clinical care, a classification system is used that categorizes patients into five classes based on thresholds for height corrected TKV (HtTKV) at a given age (A through E, with A indicating the best and E the worst prognosis with respect to future kidney function decline).[2] In addition, there is a classification indicating whether a patient is suitable for inclusion in clinical trials. This classification contains three classes: patients that should not be included in clinical trials [I], patients who's suitability should be re-evaluated at yearly intervals [II] and patients that are optimal candidates for clinical trials [III]).[2] To assess reclassification, we created 5x5 and 3x3 cross-tabulations using HtTKV limits for their specific age.[2] In these tables the proportion of reclassified participants were calculated when using HteTKV instead of HtmTKV. For this analysis only the "typical cases" were used, as advised for this classification system, defined as MRIs with cysts with bilateral and diffuse distribution, where all cysts contribute similarly to TKV.[2] Second, we assessed what the consequences were for sample size calculation for clinical trials using change in eTKV instead of change in mTKV. Sample size calculations were based on literature [13] and used data of all patients who had longitudinal follow-up data available with respect to change in mTKV and eTKV. The number of patients needed per group was calculated assuming a power of 80% and a two-sided α of 0.05 to detect a percentage difference in TKV growth between treatment groups.[14]

RESULTS

Study participants

The study population consisted of 264 patients with ADPKD. Their characteristics are listed in Table 1. These patients were relatively young with a mean age of 46.7 ± 8.8 years and showed already clear signs of disease. Most patients used anti-hypertensive medication. Their eGFR was impaired (57.0 ± 21.1 ml/min*1.73m^2), with a wide range in eGFR (from 15.2 to 129.2 ml/min*1.73m^2). Urinary albumin excretion (45.8 [21.5 - 88.7] mg/24h) and total kidney volume (1.87 [1.24 - 2.67] L) were increased.

Reproducibility of mTKV and eTKV

Table 2 shows the variability for mTKV and eTKV. The average intra-observer CV was 1.8% for mTKV and 2.6% for total liver volume, whereas the inter-observer CV was 2.3% and 3.5%, respectively. The variability for $eTKV_{ELLIPSOID}$ was significantly higher than for mTKV, whereas for $eTKV_{PANK}$ no significant differences were found when compared to mTKV. Analysis time was approximately 55 minutes per MRI for mTKV and 65 minutes for total liver volume, with higher analysis times in case of larger organs. The average time needed per MRI to estimate TKV using the mid-slice method was 15 minutes and using the ellipsoid method 5 minutes.

Performance of the TKV estimation methods

The correlations of mTKV vs. $mTKV_{REPEAT}$, $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$ are shown in Figure 1. Supplementary Figures 1 and 2 shows these correlations for left and right kidneys separately. In 38 patients no transversal scans were available, precluding estimation of TKV using the ellipsoid method, whereas in 8 patients not the whole kidneys were scanned in a coronal view, precluding estimation of TKV using the mid-slice method. High correlations were observed for all three methods: $mTKV_{REPEAT}$ r=0.998 (p<0.001), $eTKV_{ELLIPSOID}$ r=0.988 (p<0.001), and $eTKV_{PANK}$ r=0.987 (p<0.001). Figure 1 also shows Bland-Altman plots of mTKV vs. the percentage difference between mTKV and $mTKV_{REPEAT}$ and both eTKV methods. $mTKV_{REPEAT}$ showed low bias (mean -0.03% ± 3.4%). eTKV also did not systematically over- or underestimate mTKV (bias of 1.2% ± 9.3% and 4.8% ± 8.3% for $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$, respectively, Table 3). Bias for $eTKV_{PANK}$ was significantly higher than for $mTKV_{REPEAT}$ (p=0.001), whereas bias for $eTKV_{ELLIPSOID}$ did not significantly differ from $mTKV_{REPEAT}$ (p=0.7). Given the lower SD, $mTKV_{REPEAT}$ had a better precision and therefore better performance when compared to $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$.

In addition, when these analyses were repeated with ADPKD patients stratified for eGFR, we observed no significant difference in bias for $eTKV_{ELLIPSOID}$ and $mTKV_{REPEAT}$ in patients with eGFR ≥ 60 mL/min*1.73m^2 vs. eGFR <60 mL/min*1.73m^2 (p=0.9 and p=0.6, respectively). Between $eTKV_{PANK}$ and $mTKV_{REPEAT}$, we observed a significant difference in patients with eGFR < 60 mL/min*1.73m^2 (p=0.002) and no significant difference in patients with eGFR ≥60 mL/min*1.73m^2 (p=0.5) (Supplementary Table 2).

When investigating factors associated with bias, it appeared that liver volume was not associated with bias in $eTKV_{ELLIPSOID}$ (p=0.3) and $eTKV_{PANK}$ (p=0.1). Bias was also not associated with age (p=0.4 and p=0.08), height (p=0.6 and p=0.14) and strength of magnetic field (p=0.9 and p=0.3), respectively for $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$. In 35 MRI-scans, the T2-weighted images showed too low quality and therefore T1-weighted images were used. A significantly higher bias for $eTKV_{ELLIPSOID}$ was observed in T1-weighted images than in T2-weighted images (p=0.007), but not for $eTKV_{PANK}$ (p=0.9), with T1-weighted images leading to underestimation of TKV.

Table 2. Intra- and inter-observer coefficient of variation (CV) for measured total kidney volume (mTKV) and for estimated total kidney volume using the ellipsoid method ($eTKV_{ELLIPSOID}$) and the mid-slice method ($eTKV_{PANK}$).

	Left kidney	Right kidney	Both kidneys
mTKV			
Number of scans	10	10	10
Intra-observer CV (%)	2.3	1.9	1.8
Inter-observer CV (%)	2.6	2.9	2.3
$eTKV_{ELLIPSOID}$			
Number of scans	10	10	10
Intra-observer CV (%)	4.9*	4.3*	3.9*
Inter-observer CV (%)	6.0*	8.5*	6.3*
$eTKV_{PANK}$			
Number of scans	10	10	10
Intra-observer CV (%)	3.8	3.1	3.0
Inter-observer CV (%)	4.2	3.1	3.4

* p-value <0.05 for difference in intra- or inter-observer CV eTKV vs. corresponding value of mTKV.

Ability to detect changes in TKV when using estimation methods

Follow-up data for mTKV were available for 48 patients. Their baseline characteristics are given in Table 1. These patients were younger, showed less signs of disease, with a higher eGFR (79.7 ± 22.6 ml/min*1.73m^2) and lower urinary albumin excretion (46.2 [19.0 - 181.0] mg/24h) than the overall study group. During a follow-up of 3.0 years their mTKV increased from 1.79 (1.36 - 2.56) to 2.18 (1.55 - 2.73) L ($p<0.001$). The median difference during follow-up in TKV was 0.25 (0.04 - 0.54), 0.30 (0.08 - 0.86) and 0.28 (0.08 - 0.54) L for mTKV, $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$, respectively (Table 4). Change in eTKV compared to change in mTKV was not significantly different for both estimation methods ($p=0.2$ and $p=0.5$ for $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$, respectively). Figure 2 plots the percentage change in mTKV vs. the percentage change in eTKV. High concordance correlations were observed for $eTKV_{ELLIPSOID}$ ($r=0.798$, $p<0.001$) and $eTKV_{PANK}$ ($r=0.866$, $p<0.001$). Percentage change in eTKV did not show systematic under- or overestimation, with bias and precision (% change mTKV - % change eTKV) -2.2% ± 10.3% and -1.8% ± 8.3% for $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$, respectively (Figure 2). In the majority of the patients, bias for change in TKV was between -10% and 10% (72.3% and 74.5% of patients for $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$, respectively).

Consequences of using eTKV instead of mTKV

When using the eTKV methods instead of mTKV for risk classification with respect to prognosis for rapid kidney function decline, we excluded the radiologically atypical ADPKD cases (n=30), as advised for this classification system. 93.3% ($eTKV_{ELLIPSOID}$) and 90.7% ($eTKV_{PANK}$) of the patients were reclassified to their original risk categories (Table 4), whereas for both estimation methods, less than 1.5% of the patients were reclassified to a higher risk category and less than 8% to a lower risk category. For classification for selection of patients for clinical trials, we observed that 97.4% ($eTKV_{ELLIPSOID}$) and 96.9% ($eTKV_{PANK}$) of the patients were reclassified to their original categories. No patients were reclassified to a higher risk category when using $eTKV_{ELLIPSOID}$ and only 1 patient when using $eTKV_{PANK}$ (Table 5).

The consequences of using percentage change in eTKV instead of percentage change in mTKV as endpoint for sample size calculation for randomized controlled trials were assessed using data of the 48 ADPKD patients of which follow-up data were available. We calculated the number of study participants per treatment group needed to be enrolled to demonstrate a certain percentage decrease in rate of growth in TKV. The results are shown in Supplementary Table 3. To detect for instance a 30% decrease in rate of growth in mTKV over a period of 3 years, 186

patients are needed per treatment group, whereas for $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$ these numbers are 122 and 143, respectively.

Table 3. Performance of the ellipsoid method and the mid-slice method to estimate total kidney volume ($eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$, respectively).

	$eTKV_{ELLIPSOID}$	$eTKV_{PANK}$	p-value*	p-value%
Number of scans	226	256		
Left kidney volume (L)	1.03 (0.65 - 1.48)	0.93 (0.61 - 1.40)	0.3	<0.001
Bias (%)	-0.7	5.6	0.9	0.001
Precision (%)	11.8	10.0		
Right kidney volume (L)	0.90 (0.57 - 1.37)	0.82 (0.53 - 1.27)	0.003	<0.001
Bias (%)	1.7	3.8	0.2	0.03
Precision (%)	12.7	11.9		
Total kidney volume (L)	1.93 (1.25 - 2.80)	1.80 (1.17 - 2.52)	0.008	<0.001
Bias (%)	1.2	4.8	0.7	0.001
Precision (%)	9.3	8.3		
Accuracy				
P_{10}	77.6	79.4	<0.001	<0.001
P_{15}	91.9	90.2	<0.001	<0.001
P_{20}	96.9	93.7	<0.001	<0.001
CCC	0.988	0.987		

Note: P values are calculated by paired t test when normally distributed, Wilcoxon signed-rank test when non-normally distributed for continues variables and McNemar test for nominal variables. Abbreviations and definitions: Bias, mean % difference between mTKV and eTKV; Precision, 1 standard deviation of bias; Accuracy, percentage of eTKV values within 10% (P_{10}), 15% (P_{15}) and 20% (P_{20}) of their corresponding mTKV value; CCC, concordance correlation coefficient.
* $mTKV_{REPEAT}$ vs. $eTKV_{ELLIPSOID}$
% $mTKV_{REPEAT}$ vs. $eTKV_{PANK}$

Table 4. Baseline and follow-up total kidney volume (TKV) data in 48 ADPKD patients follow-up data available.

	Baseline (L)	Follow-up (L)	Change (L)	Change (%)
Left kidney				
mTKV	0.99 (0.74 – 1.39)	1.23 (0.83 – 1.56)	0.13 (0.01 – 0.29)	15.0 ± 18.7
$eTKV_{ELLIPSOID}$	1.03 (0.70 – 1.44)	1.26 (0.85 – 1.58)	0.10 (0.04 – 0.37)	17.7 ± 18.1
$eTKV_{PANK}$	0.92 (0.68 – 1.24)	1.10 (0.78 – 1.44)	0.17 (0.04 – 0.36)*	19.7 ± 19.0*
Right kidney				
mTKV	0.80 (0.57 – 1.17)	0.99 (0.68 – 1.29)	0.13 (0.06 – 0.25)	19.4 ± 18.6
$eTKV_{ELLIPSOID}$	0.81 (0.58 – 1.10)	1.04 (0.65 – 1.39)	0.14 (0.04 – 0.29)	23.1 ± 22.8
$eTKV_{PANK}$	0.78 (0.60 – 1.14)	0.90 (0.65 – 1.24)	0.13 (0.04 – 0.24)	17.0 ± 19.6
Both kidneys				
mTKV	1.79 (1.36 – 2.56)	2.18 (1.55 – 2.73)	0.25 (0.04 – 0.54)	16.7 ± 17.1
$eTKV_{ELLIPSOID}$	1.86 (1.32 – 2.75)	2.39 (1.50 – 2.80)	0.30 (0.08 – 0.86)	19.3 ± 16.1
$eTKV_{PANK}$	1.79 (1.12 – 2.43)	2.03 (1.49 – 2.63)	0.28 (0.08 – 0.54)	17.8 ± 16.1

Abbreviations: mTKV, measured total kidney volume; $eTKV_{ELLIPSOID}$, total kidney volume estimated using ellipsoid method; $eTKV_{PANK}$, total kidney volume estimated using mid-slice method.
No significant differences between change in eTKV vs. change in mTKV were noted, but only at change in left $eTKV_{PANK}$ as indicated with an asterisks. * p-value <0.05.

Table 5. Reclassification for staging into risk categories for rapid kidney function decline for clinical care (A-E) and for selection of patients for clinical trials based on thresholds for height corrected TKV at a given age (I-III) using ellipsoid method ($eTKV_{ELLIPSOID}$) and using mid-slice method ($eTKV_{PANK}$) instead of mTKV.

		$eTKV_{ELLIPSOID}$						*$eTKV_{PANK}$*				
		A	B	C	D	E		A	B	C	D	E
mTKV	A	**5**					A	**5**	1			
	B		**28**				B	1	**34**			
	C		5	**66**	2		C		6	**80**	2	
	D			5	**46**		D			7	**49**	
	E				1	**37**	E				4	**36**
		$eTKV_{ELLIPSOID}$						*$eTKV_{PANK}$*				
		I		II		III		I		II		III
mTKV	I	**5**					I	**5**		1		
	II			**28**			II	1		**34**		
	III			5		**157**	III			6		**178**

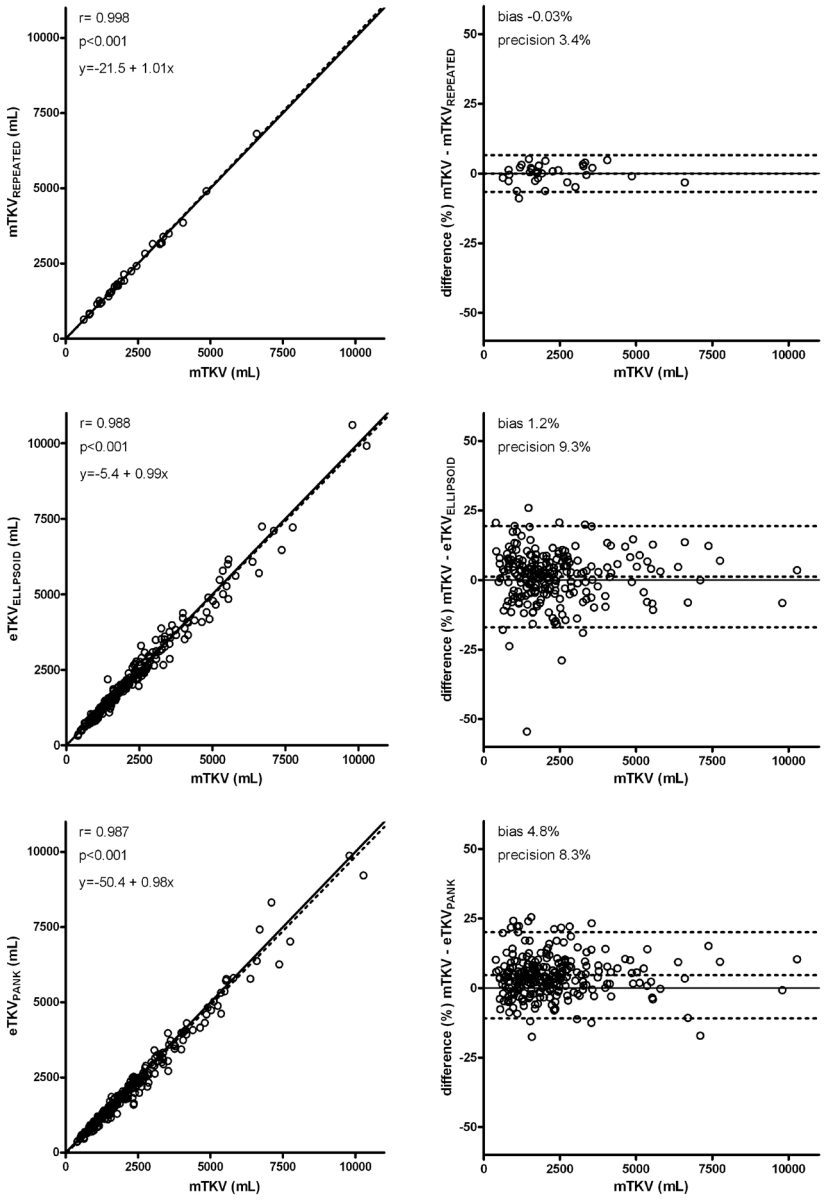


Figure 1. Associations between measured total kidney volume (mTKV) and repeated mTKV ($mTKV_{REPEAT}$) (upper panels), estimated TKV using the ellipsoid method ($eTKV_{ELLIPSOID}$) (middle pannels) and the mid-slice method ($eTKV_{PANK}$) (lower panels). Left panel shows scatter plots (solid line representing the line of identity and the dotted line the actual regression line), whereas the right panel shows Bland-Altman plots (solid line indicating no difference and dotted lines representing mean difference [i.e. bias] with 95% confidence interval).

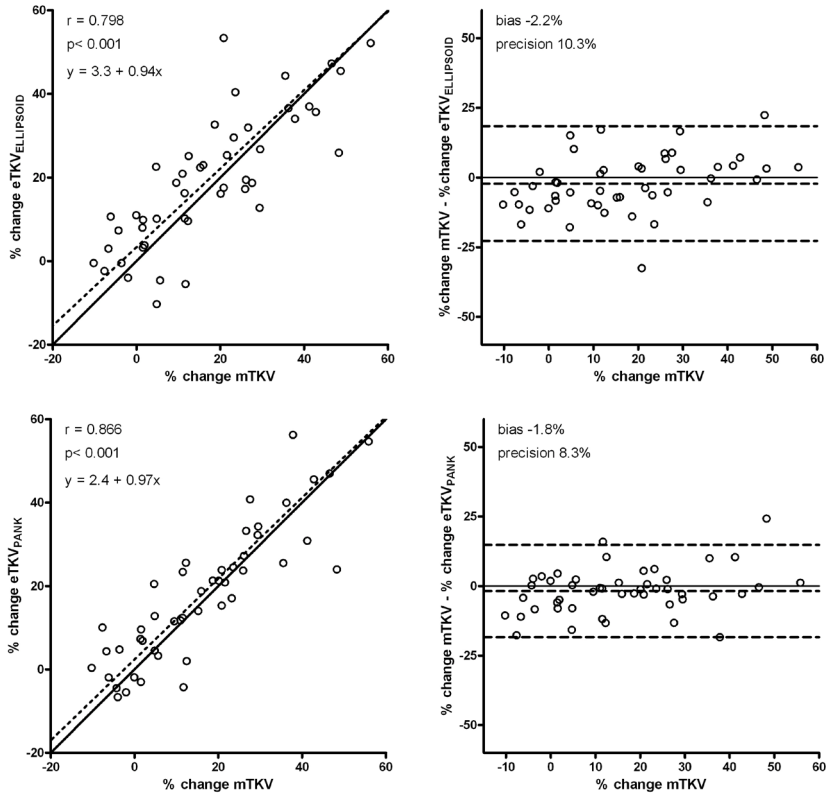


Figure 2. Associations between percentage change in measured total kidney volume (mTKV) and percentage change in estimated total kidney volume (eTKV) using the ellipsoid method and the mid-slice method in 48 ADPKD patients who had follow-up data available. Left panel shows scatter plots (solid line representing the line of identity and dotted line the actual regression line), whereas the right panel shows Bland-Altman plots (solid horizontal line indicating no difference and dotted lines representing mean difference [i.e. bias] with 95% confidence interval).

DISCUSSION

This study was conducted to investigate whether TKV can be estimated accurately using the mid-slice and ellipsoid methods in a group of ADPKD patients with a wide range of kidney function. In a test-set of ten ADPKD patients we found that both estimation methods were highly reproducible. In our study cohort of 264 ADPKD patients both methods showed low bias, high precision and high accuracy when compared to mTKV. This held for the overall cohort, as well as for patients with higher and lower eGFR. In the 48 patients who had follow-up MRIs available, change in eTKV was not different from change in mTKV for both methods.

Measurement of TKV using the gold standard method is time consuming and needs specific software, which limits its applicability for clinical care. Methods have therefore been sought to estimate TKV in a more feasible way. Two methods have been published recently,[2,7] which, however, have yet not been validated. This formed the rationale to perform the present study. For determination whether these estimation methods can be used to assess TKV, it is of importance to answer the following five questions.

First of all, it is important to investigate what the reliability of the gold-standard method is. In our study we found that the variability in stereologic volumetric assessment (i.e. the gold standard method) of TKV was very low. Our results for mTKV are in line or even better than the variability for mTKV obtained by the CRISP consortium[6] (intra-observer CV 1.9% vs. 1.8% in the present study; inter-observer CV 3.5% vs. 2.3% in the present study). Kistler et al also investigated the reproducibility of mTKV,[15] but did not report specifically intra- and inter-observer CVs and used Lin's concordance correlation coefficient (CCC) instead. Their CCC for mTKV was high, and similar to our result (CCC 1.000 vs. 0.998 in the present study). We preferentially used T2-weighted images without contrast to measure TKV, while the CRISP consortium used for their data gadolinium enhanced-contrast T1-weighted images and Kistler et al unenhanced T1-weighted images. The higher kidney-tissue contrast in the T2-weighted images helps to better delineate the kidney boundaries against the surrounding tissues when compared to T1-weighted images. Only when the T2-weighted images did not cover both kidneys we used T1-weighted images. Unenhanced T1-weighted images have lower signal and lower intrinsic tissue-contrast compared with gadolinium enhanced contrast T1-weighted images. Bae et al showed that unenhanced T1-weighted volumes were significantly lower than contrast-enhanced T1-weighted volumes.[16] These differences were more pronounced in smaller kidneys, because in such cases the ratio of kidney boundaries area to kidney volume is higher. In line with these findings, we observed a significantly higher bias for $eTKV_{ELLIPSOID}$ on unenhanced T1-weighted images when compared to unenhanced T2-weighted images. These data in combination indicate that unenhanced T1-weighted images underestimate mTKV. The use of gadolineum is nowadays discouraged in patients with chronic kidney disease, because of fear of gadolineum associated nephrotoxicity and systemic fibrosis.[17,18] Therefore, it seems advisable to measure or estimate TKV using unenhanced T2-weighted images.

Second, do these estimation methods show low variability? Both TKV estimation methods showed limited variability. Although the variability in mTKV was lowest, and significantly lower than for $eTKV_{ELLIPSOID}$, intra- and inter-observer CVs were satisfactory for $eTKV_{ELLIPSOID}$ as well as for $eTKV_{PANK}$, with slightly better values for $eTKV_{PANK}$.

Third, does the estimation method show good agreement with the gold standard method? We found for both estimation methods that eTKV correlated strongly with mTKV. Although bias and precision showed again better values for $mTKV_{REPEAT}$ (-0.03% and 3.4%, respectively), the results for $eTKV_{ELLIPSOID}$ as well as for $eTKV_{PANK}$ were good. Bias was low for $eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$, (1.2% and 4.8 %, respectively), although for $eTKV_{PANK}$ slightly (but significantly) higher than for $mTKV_{REPEAT}$. In addition, precision was reasonable, now with slightly better results for $eTKV_{ELLIPSOID}$ ($eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$ 9.3% and 8.3%, respectively, Table 3). Consequently we found good accuracy for both estimation methods ($eTKV_{PANK}$ P_{20} 93.7%, and $eTKV_{ELLIPSOID}$ P_{20} 96.9%). Our findings with respect to accuracy are consistent with the values obtained in the cohort in which the ellipsoid method was developed (P_{10} 70.3% vs. 77.9% in the present study).[2] When stratified for kidney function, our results with respect to bias suggest that the mid-slice method may be less accurate in ADPKD patients with lower kidney function, who generally have larger kidneys. Besides these statistical data, consequences for clinical care should be investigated when answering the question whether estimation methods show good agreement with the gold standard method. Irazabal et al proposed a classification system for ADPKD patients to assess their risk for rapid kidney function decline and to guide selection of patients for clinical trials.[2] This classification system uses thresholds defined on age and height corrected TKV. We investigated the percentage of patients that are reclassified when using eTKV instead of mTKV. In the classification system for risk assessment, we observed that only a limited percentage of patients was reclassified, and that these patients were especially reclassified to a lower risk category (Table 5). No fundamental differences in results were observed for the two TKV estimation methods, and only one patient was reclassified to a risk category that would preclude treatment when using $eTKV_{PANK}$ (category B).

Fourth, can the estimation method detect changes in TKV over time? As far as we are aware no study has yet investigated the performance of estimation methods to assess changes in TKV. In our analyses, we found a high concordance correlation between change in mTKV and change in $eTKV_{PANK}$ and $eTKV_{ELLIPSOID}$ during three

years follow-up, and no difference between change in mTKV and change in $eTKV_{PANK}$ and $eTKV_{ELLIPSOID}$ (Table 4). Consequently, when data are used of change in eTKV instead of change in mTKV, similar numbers of patients have to be included in clinical trials to be able to show a decrease in rate of growth in TKV (Supplementary Table 3). These longitudinal results may seem surprising, because they appear to be in contrast with our cross-sectional data, where we showed that mTKV shows better reliability than $eTKV_{PANK}$ and $eTKV_{ELLIPSOID}$, albeit that these differences were small. In our opinion, this may be due to two explanations. It could well be that with the TKV estimation methods a systematic error is made in an individual patient in assessing TKV at baseline, for instance due to a peculiar shape of a cystic kidney, but that the same error is made during follow-up, because the shape of the cystic kidney has not changed. In this way a systematic error in baseline TKV estimation will not translate in bias in change in estimated TKV during follow-up on a patient level. In addition, the natural variability in growth in TKV between patients may be that high, that the limited variability that is added by using estimated TKV is not relevant when assessing mean change in TKV on a group level.

The fifth and last question to be answered is whether the estimation method is feasible from a clinical point of view. To estimate TKV using the mid-slice method, special software is necessary to measure the mid-slice area, limiting clinical applicability. In contrast, all clinicians can estimate TKV by the ellipsoid method using standard MRIs without special software. Furthermore, the ellipsoid method requires less time to estimate TKV than using the mid-slice method, and both methods require far less time than measurement of TKV with the gold standard method, i.e. stereologic method.

The answers to the above questions indicate that, although estimated TKV may be slightly less precise than measurement of TKV using stereologic method, it can be used with confidence in clinical care. Because the two TKV estimation methods show numerically hardly any differences with respect to bias, precision and accuracy, and no difference in ability to detect changes in eTKV, the more feasible ellipsoid method is to be preferred over the mid-slice method. Whether this conclusion is also valid for the use of $eTKV_{ELLIPSOID}$ instead of mTKV for clinical trials needs confirmation in independent, large datasets with MRIs obtained at baseline as well as during follow-up. Our data form the rationale to perform such studies.

A limitation of the present study is that our results hold primarily true for the cross-sectional correlation between mTKV and eTKV. Our results for follow-up data should

be interpreted with caution, because the results are based on a limited number of patients. Strengths of this study are that we investigated both estimation methods in a group of ADPKD patients with relatively well-preserved as well as impaired kidney function, and that we are the first to externally validate both estimation methods.

In conclusion, we demonstrated that both methods to estimate TKV perform relatively well in ADPKD patients overall, as well as in patients with preserved as well as impaired kidney function. In addition, both estimation methods detect relatively accurate changes in TKV over time. Because of these results and the higher feasibility, we advise to use the ellipsoid method for TKV estimation in clinical care. Whether this method can also be used for clinical trials deserves further study.

ACKNOWLEDGMENTS

We acknowledge R.L. Kadijk for assistance at the outpatient clinic, L. Schepel, J. van Everdink, S. Voorrips, C. Plate, I.L. van Manen, M.B.R. Wiertz, B. van der Slik and R.R. Buiten for measuring total kidney volumes and J. Grozema and A. Siebeijn-Kuiper for assistance during MR-imaging. We thank the patients involved in this study for their participation.

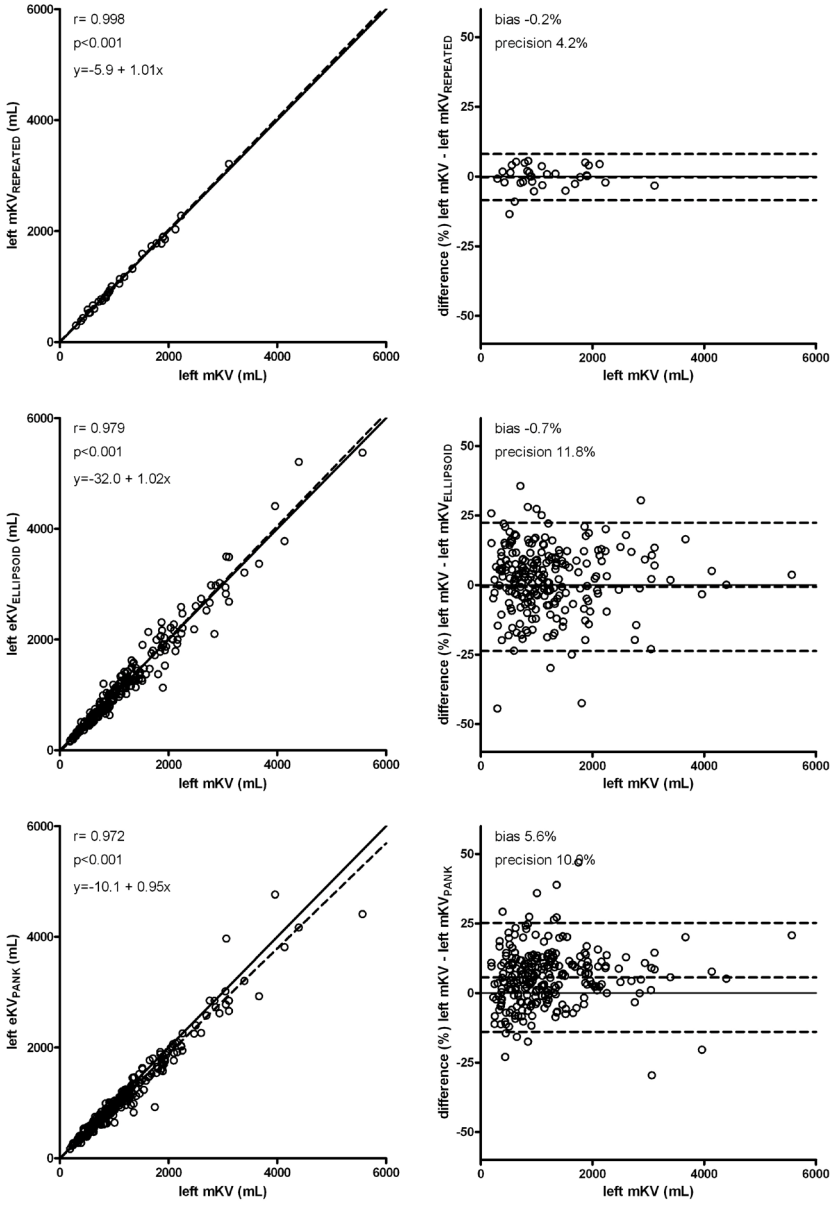
DISCLOSURES

None.

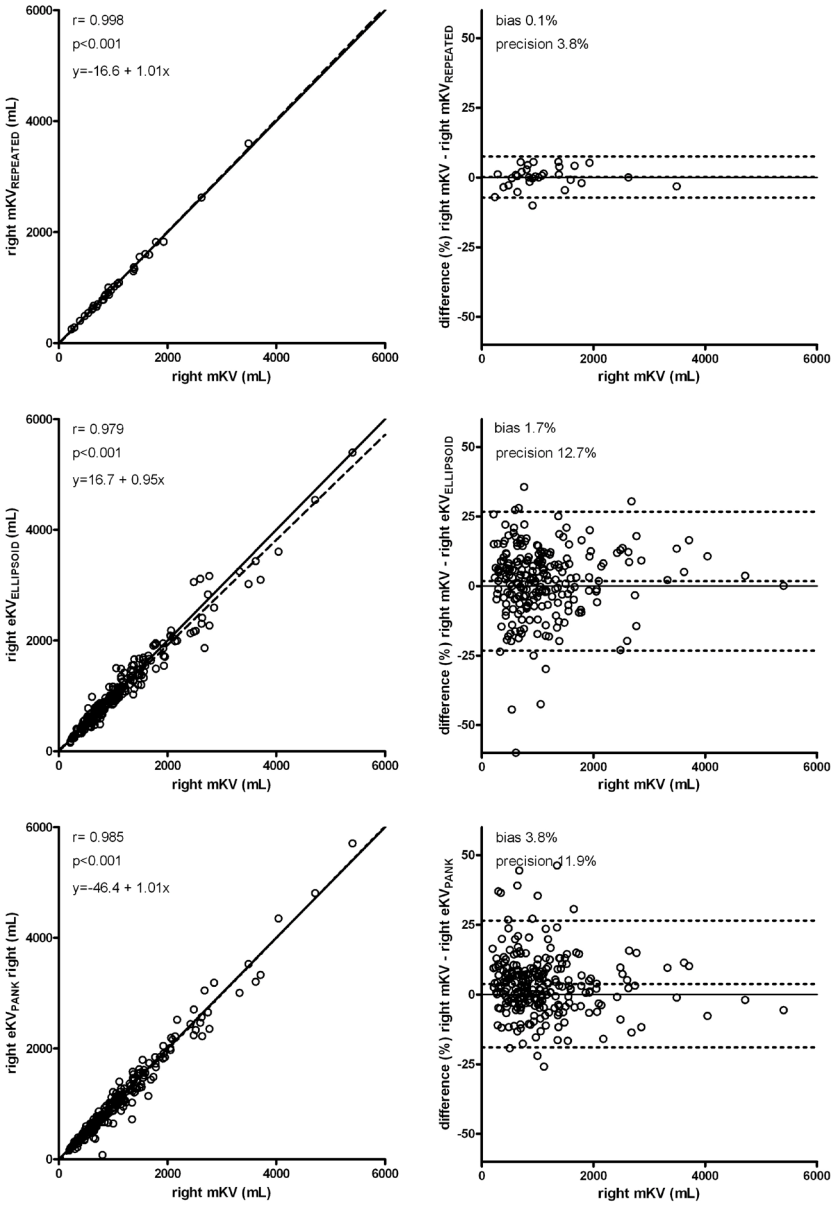
REFERENCES

1. Grantham JJ, Torres VE, Chapman AB, et al. Volume progression in polycystic kidney disease. N Engl J Med. 2006;354(20):2122-2130.
2. Irazabal MV, Rangel LJ, Bergstralh EJ, et al. Imaging classification of autosomal dominant polycystic kidney disease: A simple model for selecting patients for clinical trials. J Am Soc Nephrol. 2014.
3. Randomized Controlled Trial of Triptolide-Containing Formulation for Autosomal Dominant Polycystic Kidney Disease (ADPKD). ClinicalTrials.gov identifier: NCT02115659. . 2014.
4. Meijer E, Drenth JP, d'Agnolo H, et al. Rationale and design of the DIPAK 1 study: A randomized controlled clinical trial assessing the efficacy of lanreotide to halt disease progression in autosomal dominant polycystic kidney disease. Am J Kidney Dis. 2013.
5. Chapman AB, Torres VE, Perrone RD, et al. The HALT polycystic kidney disease trials: Design and implementation. Clin J Am Soc Nephrol. 2010;5(1):102-109.
6. Chapman AB, Guay-Woodford LM, Grantham JJ, et al. Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): The consortium for radiologic imaging studies of polycystic kidney disease (CRISP) cohort. Kidney Int. 2003;64(3):1035-1045.
7. Bae KT, Tao C, Wang J, et al. Novel approach to estimate kidney and cyst volumes using mid-slice magnetic resonance images in polycystic kidney disease. Am J Nephrol. 2013;38(4):333-341.
8. Torres VE, Chapman AB, Devuyst O, et al. Tolvaptan in patients with autosomal dominant polycystic kidney disease. N Engl J Med. 2012;367(25):2407-2418.
9. Boertien WE, Meijer E, de Jong PE, et al. Short-term renal hemodynamic effects of tolvaptan in subjects with autosomal dominant polycystic kidney disease at various stages of chronic kidney disease. Kidney Int. 2013;84(6):1278-1286.
10. Pei Y, Obaji J, Dupuis A, et al. Unified criteria for ultrasonographic diagnosis of ADPKD. J Am Soc Nephrol. 2009;20(1):205-212.
11. Levey AS, Greene T, Kusek J, Beck GJ. A simplified equation to predict glomerular filtration rate for serum creatinine [abstract]. . 2000;11:115A.
12. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604-612.
13. Noordzij M, Dekker FW, Zoccali C, Jager KJ. Sample size calculations. Nephron Clin Pract. 2011;118(4):c319-23.
14. Sample size calculator. date used 14/08/2014:http://www.stat.ubc.ca/~rollin/stats/ssize/.
15. Kistler AD, Poster D, Krauer F, et al. Increases in kidney volume in autosomal dominant polycystic kidney disease can be detected within 6 months. Kidney Int. 2009;75(2):235-241.
16. Bae KT, Tao C, Zhu F, et al. MRI-based kidney volume measurements in ADPKD: Reliability and effect of gadolinium enhancement. Clin J Am Soc Nephrol. 2009;4(4):719-725.
17. Ergun I, Keven K, Uruc I, et al. The safety of gadolinium in patients with stage 3 and 4 renal failure. Nephrol Dial Transplant. 2006;21(3):697-700.
18. Bernstein EJ, Schmidt-Lauber C, Kay J. Nephrogenic systemic fibrosis: A systemic fibrosing disease resulting from gadolinium exposure. Best Pract Res Clin Rheumatol. 2012;26(4):489-503.
19. Penfield JG, Reilly RF,Jr. What nephrologists need to know about gadolinium. Nat Clin Pract Nephrol. 2007;3(12):654-668.

SUPPLEMENTARY MATERIAL



Supplementary Figure 1. Associations between measured left kidney volume (mKV) and repeated mKV (eKV_{REPEAT}) (upper panels), estimated left kidney volume using ellipsoid method ($eKV_{ELLIPSOID}$) (middle panels) and using mid-slice method (eKV_{PANK}) (lower panels). Left panel shows scatter plots (solid line representing the line of identity and the dotted line the actual regression line), whereas the right panel shows Bland-Altman plots (solid line indicating no difference and dotted lines representing mean difference [i.e. bias] and 95% confidence interval).



Supplementary Figure 2. Associations between measured right kidney volume (mKV) and right repeated mKV (eKV_{REPEAT}) (upper panels), estimated right kidney volume using ellipsoid method ($eKV_{ELLIPSOID}$) (middle panels) and using mid-slice method (eKV_{PANK}) (lower panels). Left panel shows scatter plots (solid line representing the line of identity and the dotted line the actual regression line), whereas the right panel shows Bland-Altman plots (solid line indicating no difference and dotted lines representing mean difference [i.e. bias] and 95% confidence interval).

Supplementary Table 1. Baseline characteristics stratified by eGFR.

	eGFR ml/min*1.73m^2		p-value
	< 60	≥ 60	
N	171	92	
Age (y)	49.5 ± 7.7	41.5 ± 8.6	<0.001
Male (%)	49.4	53.3	0.5
Body mass index (kg/m^2)	26.9 ± 4.1	26.3 ± 4.7	0.2
Body surface area (m^2)	2.01 ± 0.23	2.02 ± 0.23	0.8
Diastolic blood pressure (mm Hg)	82.1 ± 9.8	80.3 ± 8.8	0.1
Systolic blood pressure (mm Hg)	133.1 ± 12.9	129.4 ± 13.0	0.03
Antihypertensive medication (%)	88.1	78.3	0.03
Plasma creatinine (mmol/L)	144.9 ± 40.7	92.7 ± 17.2	<0.001
eGFR (ml/min/1.73m^2)	44.7 ± 9.5	79.6 ± 18.1	<0.001
24h Urine volume (L)	2.44 ± 0.76	2.27 ± 0.79	0.09
Albuminuria (mg/24h)	52.5 (25.2 - 104.5)	33.4 (17.8 - 80.3)	0.03
Total kidney volume (L)	2.01 (1.39 - 3.04)	1.67 (1.12 - 2.39)	0.01
Left kidney volume (L)	1.04 (0.67 - 1.57)	0.91 (0.60 - 1.30)	0.07
Right kidney volume (L)	1.00 (0.64 - 1.43)	0.71 (0.52 - 1.05)	0.001

Abbreviations: eGFR, estimated glomerular filtration rate (CKD-EPI equation). P-values indicate differences between eGFR < 60 and ≥ 60. P-values are calculated by t-test when normally distributed and by Mann-Whitney U test when non-normally distributed.

Supplementary Table 2. Performance of the ellipsoid method and the mid-slice method to estimate total kidney volume ($eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$, respectively), stratified for eGFR ≥60 en <60 ml/min*1.73m^2.

	$eTKV_{ELLIPSOID}$			$eTKV_{PANK}$		
eGFR	≥60	<60	p	≥60	<60	p
N	76	150		92	164	
Left kidney volume (L)	0.96 (0.60 - 1.27)	1.11 (0.69 - 1.76)	0.1	0.88 (0.59 - 1.23)*	1.02 (0.65 - 1.47)*	0.1
Bias	-0.9	-0.6	1.0	4.5	6.2*	0.6
Precision	12.1	11.6		11.2	9.2	
Right kidney volume (L)	0.71 (0.52 - 1.09)	0.99 (0.63 - 1.49)*	0.005	0.72 (0.49 - 1.08)*	0.95 (0.59 - 1.38)	0.02
Bias	0.6	2.3	0.3	0.9	5.4*	0.002
Precision	12.2	13.0		9.2	12.9	
Total kidney volume (L)	1.75 (1.17 - 2.38)	2.12 (1.14 - 3.12)*	0.02	1.64 (1.11 - 2.33)*	1.90 (1.26 - 2.87)*	0.04
Bias	0.7	1.5	0.5	3.0	5.8*	0.02
Precision	8.8	9.5		7.2	8.7	
Accuracy						
P_{10}	79.7	76.5	0.6	87.9	74.7	0.03
P_{15}	91.9	91.9	1.0	92.4	88.9	0.4
P_{20}	98.6	96.0	1.0	95.7	92.6	1.0
CCC	0.986	0.988		0.989	0.985	

Note: p-values are calculated with independent t-tests when normally distributed and with Wilcoxon signed-rank tests when non-normally distributed for unpaired data, and with paired t-tests and McNemar tests for paired data. Abbreviations and definitions: $eTKV_{ELLIPSOID}$, estimated total kidney volume using ellipsoid method; $eTKV_{PANK}$, estimated total kidney volume using mid-slice method; eGFR, estimated glomerular filtration rate. Accuracy, percentage of estimated total kidney volume values within 10% (P_{10}), 15% (P_{15}) and 20% (P_{20}) of their corresponding measured total kidney volume value (TKV). Bias, mean % difference between mTKV and eTKV. Precision, 1 standard deviation of bias; CCC, concordance correlation coefficient. P values for $eTKV_{ELLIPSOID}$ ≥60 vs. <60 are calculated by t-test when normally distributed and Mann-Whitney U test when non-normally distributed. * paired t-test for $mTKV_{REPEAT}$ vs. eTKV.

Supplementary Table 3. Numbers of participants per treatment group needed for randomized controlled trials to be able to show a specific % difference in growth in total kidney volume over a period of three years when using gold standard total kidney volume (mTKV) or estimated kidney volume using the ellipsoid method ($eTKV_{ELLIPSOID}$) or mid-slice method ($eTKV_{PANK}$).kidney volume ($eTKV_{ELLIPSOID}$ and $eTKV_{PANK}$, respectively), stratified for eGFR ≥60 en <60 ml/min*1.73m^2.

	mTKV	$eTKV_{ELLIPSOID}$	$eTKV_{PANK}$
20%	417	274	332
30%	186	122	143
40%	105	69	81
50%	67	44	52

SUPPLEMENTARY METHODS

Magnetic resonance imaging

The UMC Groningen used a 1.5-Tesla MR scanner (Magneto Avanto, Siemens, Erlangen, Germany) and a 3-Tesla research MR scanner (Intera, Philips, Eindhoven, the Netherlands). All other centers used a 1.5-Tesla MRI-scan [UMC Leiden: Philips Healthcare, Eindhoven, the Netherlands; UMC Rotterdam: GE Medical Systems, Buckinghamshire, United Kingdom; and the UMC Nijmegen: Avanto Siemens, Erlangen, Germany]. Coils were placed onto the anterior and posterior abdominal walls directly over the kidneys. A short scout was scanned to localize the kidneys. Subsequently four series of images were scanned. Two T2-fast multislice spoiled gradient echo were scanned coronal and transversal, with slice thickness of 4 mm, gap/spacing 0 mm, FOV 35 cm, matrix 256*256, TE 2 ms, TR 7 ms, Flip Angle 40-50°. Thereafter a T2-single shot fast spin echo was scanned coronal (same characteristics, but different TR's and TE's per brand MRI-scanner: TE 100 ms for Siemens, TE 190 ms and TR max. 1400 ms for GE and ≈ 70 ms and TR max. 1900 ms for Philips) and a T1-3D spoiled gradient echo coronal (same characteristics except TR 4 ms and Flip Angle ≤15°). At the beginning and the end of the scan sequence had to be at least 1 slice not containing liver and kidney tissue. When a 35 cm FOV was insufficient, the FOV could be increased. Preferably, both kidneys as well as the liver, including all cysts, had to be covered within one sequence of images. When such a sequence could not be scanned, two separate sequences for liver and kidneys were allowed. The obtained MR images were anonymized and sent via a secured server to the central reading facility at the UMC Groningen, where kidney and liver volume were measured. Nine medical students were specifically trained to measure TKV. During their training period, they measured 40 kidney volumes and 20 liver volumes under supervision and guidance of an experienced MRI-technician using a standard operating procedure. After these students completed their training, they were allowed to measure TKV.

6